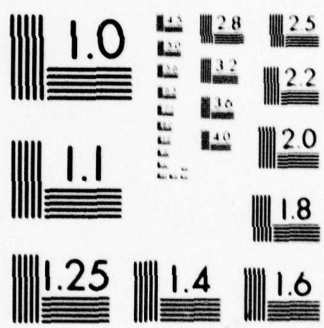
AD-A072 270 HUMAN RESOURCES RESEARCH ORGANIZATION ALEXANDRIA VA F/G 15/7
SAGEBRUSH MANEUVER TESTS: APPLIED AND THEORETICAL.(U)
MAY 56 E L SHRIVER DA-49-106-QM-1

UNCLASSIFIED NL

1 OF 1
AD
A072270

END
DATE
FILMED
9-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) SAGEBRUSH MANEUVER TESTS: APPLIED AND THEORETICAL,		5. TYPE OF REPORT & PERIOD COVERED (9) Staff Memorandum
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Edgar L. Shriver		8. CONTRACT OR GRANT NUMBER(s) (15) DA-49-106-qm-1
9. PERFORMING ORGANIZATION NAME AND ADDRESS Human Resources Research Organization (HumRRO) 300 N. Washington Street Alexandria, Virginia 22314		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Department of the Army Washington, D.C.		12. REPORT DATE (11) May 1956
		13. NUMBER OF PAGES 69
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) (12) 80p.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Research performed by HumRRO Training Methods Division under Project MANEVAL.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

measuring techniques
atomic field army (ATFA)
tests
maneuvers

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper is concerned with the problem of selecting, by scientific methods, the Army organizations and doctrines which will best complement newly developed weapons and equipment. This paper reports the trials, errors and successes of the Army Maneuver Test and Evaluation Group (AMTEG) in their work of testing and evaluating the organization and doctrine employed during the SAGEBRUSH maneuver.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

405 260

JB

Task: MANEVAL

Staff Memorandum

SAGEBRUSH MANEUVER TESTS: APPLIED AND THEORETICAL

Edgar L. Shriver

May 1956

Training Methods Division

Approved W. A. McClelland
William A. McClelland
Director of Research

HUMAN RESOURCES RESEARCH OFFICE
The George Washington University
operating under contract with
The Department of the Army

BRIEF

This paper is concerned with the problem of selecting, by scientific methods, the Army organisations and doctrines which will best complement newly developed weapons and equipment. One vehicle for such scientific selection is the Army maneuver. This paper reports the trials, errors and successes of the Army Maneuver Test and Evaluation Group (AMTEG) in their work of testing and evaluating the organisation and doctrine employed during the SAGEBRUSH maneuver. The activities associated with the Human Resources Research Office advisor to the AMTEG are reported in particular detail.

The last section of this paper is a theoretical discussion of scientific considerations regarding future maneuver tests and the development of an integrated approach to the general problem of selecting Army organisations and doctrines through the use of scientific methods.

TABLE OF CONTENTS

Page

Page

Part 1

THE TESTING PROGRAM IN THE SAGEBRUSH MANEUVERS

I. Introduction

A. Scope and Purpose

The Human Resources Research Office was invited to assign scientific advisers to Army groups responsible for testing the organization and doctrine of the new atomic field army (ATFA).[1/] These tests were conducted during the FOLLOW ME, BLUEBOLT II, and SAGEBRUSH maneuvers in 1954 and 1955.

The purpose of this paper is to describe and/or analyze the problems of constructing an empirical test of the organization and doctrine under which an Army unit of division or larger size operates. Although this memorandum is concerned primarily with the largest of the maneuvers, SAGEBRUSH, the writer believes that the text may be generalized to the other two as well.

The scientific advice given by the HumRRO representative did not deal primarily with measuring techniques such as replication, questionnaire development, judgmental scales and so on; it was more general in nature. The level of abstraction of the variables to be measured was one of the central problems the Army faced in constructing the SAGEBRUSH test. In the advice given in this area, the general scientific approach of identifying variables which predict from the abstract to the specific was applied.

This report describes the "false starts" with regard to the level of abstraction of the variables to be measured during SAGEBRUSH,

---

1/ The writer was not able to find an authoritative source that explained the meaning of the term ATFA. It was variously reported as a contraction of atomic field army and A-type field army.

recounts the counsel given by the HumRRO scientific adviser on this point, and discusses the level of abstraction that should be used in the future for predictions to be manageable yet accurate. Also described are the requirements, in terms of scientific and military effort, that must be fulfilled in order to replace "artistic prophecies" with accurate, scientific predictions of the effectiveness of various field-army organizations and doctrines.

B. Approach Adopted for SAGEBRUSH Testing

One central aspect of the approach to this problem must be specified. The approach adopted was that the test would have to be conducted in a situation as similar to combat as possible. Although tests can be and often are conducted in a much more abstract setting, the approach taken for this test is common to most others that are adopted in the initial scientific efforts in any new area.

The approach necessitated the construction of a test criterion for the ATFA field army. This requirement was fulfilled, in general form, by the ATFA field army performing in SAGEBRUSH. This maneuver was an abstraction from actual combat. The objective of the testing program was to keep the measurements of this field army at as low a level of abstraction as possible within the administrative requirements then in existence.

Parenthetically, it should be said that the writer does not assume that a duplication of combat such as SAGEBRUSH is necessary in every future test situation. It was adopted in this case because at present too little is known about the effects of omitting

the various aspects of combat. As more is learned about the importance of various aspects of the field army, the peripheral aspects can be omitted.

The scientific advice given to the Army Maneuver Test and Evaluation Group (AMTEG) concerned the abstractions from combat that had to be made in constructing the test criterion. This was a matter of defining the minimum amount of abstraction needed for the situation that existed. The advice was definitely <u>not</u> in the form of <u>what</u> to abstract, it dealt with <u>how much</u> to abstract and what techniques to use in the abstracting.

"Abstractive process" is a generic term for such activities as a job analysis, task analysis, activity analysis, test criterion building, or item development. As used by behavioral scientists, it is a process not limited to formal areas such as training, selection, and leadership. It is the process underlying the discussion in this paper.

C. <u>Historical Introduction</u>

1. <u>Maneuvers, War Games, and Command Post Exercises</u>

Large-scale tests of Army organization and doctrine have been conducted during four maneuvers, TRIANGULAR DIVISION (1939, FOLLOW ME (1954), BLUEBOLT (1954) and SAGEBRUSH (1955). Maneuvers are a particular kind of Army exercise, as are war games and command post exercises (CPX's). The following rule of thumb will serve for distinguishing among the three types:

War games are conducted "on paper." No troops actually appear in the field; they are simulated by unit symbols moved on a

terrain map. The participants in war games take the roles of commanders and move these units in accordance with their plan of battle. The degree of simulation of movement time, plans, casualty assessment, and similar factors varies with the specific war games.

CPX's require participants in the field. The participants take their position at their appropriate command posts, then plan and issue normal instructions for the conduct of the battle. Their plans are not executed by troops. Umpires take the role of units and execute the commands on paper. The umpires return information to the participants according to their diagnosis of what would have happened if the orders had been actually executed by troops in the field.

Maneuvers require both commanders and troops in the field. The commands are executed by actual units, which communicate with the commander in the usual combat fashion. Umpires are utilized for casualty assessment and for administrative decisions.

Maneuvers may utilize groups as small as a battalion or as large as one or more field armies. They may be conducted by a technical service, combat army, or individual unit, or by CONARC. When maneuvers of all types are considered, the rate of occurrence is several per year. The primary purpose is training; the maneuver may, of course, bring to light inadequacies in organization or equipment, but this is a by-product of the training aspect. Maneuvers involving very large commands, such as field armies, simulate portions of this force by special umpires, as in a CPX.

2. <u>The SAGEBRUSH Maneuver</u>

SAGEBRUSH involved two opposing 15-division field armies. The equivalent of about five divisions was actually present; the remaining divisions were simulated by special umpire groups. An area of over seven million acres was occupied by the divisions physically present.

The troops were almost evenly divided between the "Aggressor" side and the U. S. side. The maneuver was divided into four tactical phases: the retrograde movement of the U. S. field army, the static phase while the U. S. side built up its forces, the river crossing of attacking U. S. forces, and the overrunning of the Aggressor army by the U. S. army. The front lines moved almost 100 miles during the U. S. retrograde movement. A number of types of nuclear tactical weapons (simulated) were used by both sides during the maneuver.

The following excerpt from a newspaper article gives an index of the realism of the SAGEBRUSH maneuver, as well as a brief account of the action and the use of atomic weapons.

> The Aggressor commander gave the main attack job to the 82nd Airborne Division. Spread over a 17,000-yard instead of the normal 10,000-yard front, the crack paratroopers marched northward within minutes after Aggressor air launched the war. Attacking United States roadblocks from the rear, they moved rapidly forward.
>
> The Agressor commander sent the 11th Armored Cavalry Regiment (light armor) against the enemy's center in a more leisurely advance. Simultaneously, a reinforced infantry regiment attacked on the right flank, following up an initial 280-mm. atomic cannon assault on a concentration of United States troops at a road network on the right.

The Aggressor commander's aim seemed to be a double envelopment of the United States 3d Infantry Division. The defending ground commander had entrenched in high ground north of this post. If the Aggressor could compress the defenders into a small area bounded by lakes and swamps, they would offer a good target for an atomic weapon.

Meanwhile, the Aggressor commander brought up his Fourth Armored Division from reserve; if an atomic blow was successful in the "kill zone," the armor would be sent through the gap to exploit the victory. Later, the 82nd and the armor could press forward rapidly, perhaps aided by paratroop drops to disorganize the defenders.

The obvious Aggressor strategy also included bombing out the bridges over the Red River to trap United States forces south of the stream. They then would have to attempt a river crossing over temporary bridges and would again present a profitable atomic target.

The U. S. side's obvious defense tactics were to stick closely enough with his retreating infantry to the Aggressor forces to avoid becoming vulnerable to atomic attack; and by counter-attacks with his First Armored Division to concentrate the enemy as targets for his 77th Special Forces Group, possessing the 280-mm. atomic cannon, the Corporal atomic guided missile and the Honest John rocket.

3. Testing During Maneuvers

Testing of organization and doctrines was a major mission of the four maneuvers mentioned, though not as important as training. Atomic capabilities have necessitated theoretical changes in the organization and doctrine of the field army, and these changes were tested during the FOLLOW ME, BLUEBOLT, and SAGEBRUSH maneuvers. Further theoretical changes are already in existence for future field armies; it is reasonable to assume that they also will be tested in maneuvers, CPX's, or war games in the near future.

The testing duties in SAGEBRUSH were divided among several agencies, all responsible to the central testing agency, the Army Maneuver Testing and Evaluation Group. During SAGEBRUSH, the BLUEBOLT and FOLLOW ME evaluation groups continued tests (begun in 1954) at the division level. The AMTEG conducted tests at a field army level and below (to division) in addition to their nominal supervision of all testing agencies. Other testing agencies included the Combat Operations Research Group, Office of Special Weapons Development, and Project Michigan.

Operation SAGEBRUSH was not originally planned as a test vehicle. The AMTEG was not established until several months after the planning for the maneuver was started, and this had the important effect of limiting the AMTEG to the use of testing techniques which did not require any changes in operational plans. Thus, the testing control that could be imposed was severely reduced.

4. HumRRO Participation in Maneuver Tests

This discussion has been narrowed to the type of maneuvers in which organization and doctrine are materially changed on a large scale, and in which some kind of empirical measurement of the desirability of these changes is required. The history of such tests is short and HumRRO's history of participation is even shorter. HumRRO was requested to participate in the FOLLOW ME tests, as was the Operations Research Office (ORO), and both organizations assigned representatives to the testing group staff of that maneuver.

The ORO participation developed into a special test of one aspect of the over-all FOLLOW ME test, which they designed and conducted primarily with their own personnel. HumRRO's participation took the form of advice for designing the over-all testing program. HumRRO's advice was practical rather than theoretical; the administrative difficulties were accepted as inherent in the tests. The testing situation was far from ideal, and the advice was limited to what could be accomplished under the existing conditions.

HumRRO participated at both the BLUEBOLT and AMTEG echelons in an advisory capacity during SAGEBRUSH. The HumRRO adviser also served as a primary source of continuity between the FOLLOW ME testing program and the one developed by AMTEG for use in SAGEBRUSH.

The degree of change in organization and doctrine was much greater in SAGEBRUSH in the levels above division than had been the case in the BLUEBOLT and FOLLOW ME division-size maneuvers. Below division level, fairly specific TO&E changes were made; above division level, fairly large changes in concept were made. One of the primary changes was the "splitting off" of the logistics system from the operations system above the division level in the atomic field army.1/ In contrast to the FOLLOW ME and BLUEBOLT tests, specific recommendations for changes in TO&E structure were not required from the SAGEBRUSH maneuver; the recommendations from that maneuver were to be more general. An evaluation of the "separation" concept rather than the details of its implementation was desired. This meant that the nature of the testing program would be different from the FOLLOW ME program.

1/ For further orientation of the reader, this subject is discussed later, under the heading "The Principal ATFA concept."

This difference was not fully appreciated when the AMTEG first started its work, and it resulted in the development initially of a program that would have collected data too detailed to be of practical use. When this approach was seen to be impractical, the planners reverted to the opposite extreme; they began constructing a test which would produce highly abstract data. The final test was such that the data collected were somewhere between these two extremes. This subject and the HumRRO adviser's role are discussed in greater detail in section III.

This is a brief history of Army theory testing programs and the part HumRRO has played in their development. Currently there are additional theories designed to utilize the technological advances in weapons and equipment to greatest advantage. These theories are designed in sufficient detail to reorganize the Army. As the "hardware" on which they are based comes into production, new Army organizations and doctrines will probably be tested and/or adopted.

HumRRO has recommended, in the HumRRO Annex to the AMTEG report,[1] that future testing effort be focused on controlling contamination factors by manipulation of the maneuver conditions, rather than on further refinement of the present testing techniques. HumRRO has further recommended that, in order to accomplish this, a testing agency be established several months before the operational planning for the maneuver begins, at CONARC or Department of the Army level. This will allow the testing group time to make plans and recommend maneuver conditions suitable to the testing program adopted.

1. Report of Army Tests, Exercise SAGEBRUSH (CONFIDENTIAL, MODIFIED HANDLING). 4th Army Headquarters, February 1956.

D. The SAGEBRUSH Testing Program

1. The Testing Organization

The SAGEBRUSH maneuver was originally designed to train troops in the new ATFA organization and doctrine. Several months after the planning for the maneuver was initiated, it was decided also to test the ATFA concepts during the maneuver. A testing agency (AMTEG) was established. It consisted of about 20 officers, primarily majors and lieutenant colonels, and one civilian scientific adviser from HumRRO; the chief of the group was a colonel. This group will be referred to hereafter as the nucleus group. The AMTEG was concerned with testing the ATFA concepts at army, corps, and division level.

Two major subsidiary testing groups, which had been in existence the previous year during the FOLLOW ME and BLUEBOLT division-sized tests, tested the infantry and armored divisions.

Shortly before the maneuver took place a new chief and deputy chief were appointed, a brigadier general and a colonel. Shortly after these changes the nucleus group was augmented by about 80 officers, primarily lieutenant colonels and colonels, who functioned as field evaluators. After the maneuver the group was reduced to about 25 officers, most of whom had been in the original nucleus group. The additional officers were of higher rank than those of the original group and were named as the responsible officers in preparing the report of the test. When the final report was finished, two months after the maneuver, the AMTEG was disbanded. The report was submitted to CONARC with copies to the various Army schools and HumRRO for comment.

2. <u>The ATFA Concepts</u>

The ATFA concepts have never been defined exactly. Many concepts were tried for the first time during SAGEBRUSH; most of them were advanced as "desirable" in atomic war and thus categorized as "ATFA". It was the responsibility of AMTEG to test all new concepts, but some of the concepts applied to the present standard-type army as well as the ATFA army. For instance, the accelerated data-processing concept (introduced by Project Michigan), the integrated intelligence system, and the Signal Corps grid communications system apply equally well to the ATFA and to standard armies. Concepts that did not apply to the standard army were the support command (a specialized logistics concept), the tactical "Islands of Defense," and, to a certain extent, the extended ground reconnaissance concept.

The AMTEG was not primarily concerned with testing the tactical concepts. The principal concern of its testing program was the support command concept. This is generally the concept to which the term ATFA refers, although any or all of the other concepts tested during SAGEBRUSH may be included in a loose sense. (The AMTEG at one time planned to prepare more exact definitions of what constituted "ATFA", but decided against this action when differences of opinion developed regarding what "ATFA" was and what it was not. Since the AMTEG's mission was to test all new concepts, it was not crucial for them to make the distinctions for their report.)

3. The Principal ATFA Concept

The purpose of the support command concept is to relieve the tactical commander of the logistical load, and also to facilitate the flow of supplies by establishing a specialized branch for this purpose. The concept of specialisation for this function is not entirely new; it exists to some extent in the standard army, although it is not as formal or elaborate as in the ATFA army.

In the ATFA army a general officer is in command of a logistics organization responsible to him rather than to tactical commanders at various echelons; all command that tactical commanders below the army level exercised over logistical matters in the standard army is severed. The support command commander is responsible to the army (tactical) commander, but not to any of the latter's subordinates. In the ATFA army, all command previously exercised by tactical commanders over their support takes the form of requests. This ATFA concept is designed to create a greater flexibility as well as to provide other advantages. The flexibility is present in the sense that the centralised support of the logistics command may be directed toward the tactical unit in greatest need of support.

The opposite extreme is for each tactical unit to have its own support. This type of organization is inflexible, in that support cannot easily be shifted to the most needy tactical unit. The support command concept is similar to the artillery concept, according to which all the centralised artillery support can be shifted to the most needy unit.

This brief description of the support command concept does not indicate how the concept is implemented,[1] and does not point out the undesirable aspects. It only describes the general nature of the concept and what it is designed to accomplish.

E. The Abstractive Process in Army Tests

1. HumRRO Advice During SAGEBRUSH

Although HumRRO was invited to give scientific advice on test construction during the FOLLOW ME, BLUEBOLT II, and SAGEBRUSH maneuvers, this was not, strictly speaking, a HumRRO activity in that it was not in the area of either training or motivation, morale, and leadership. However, HumRRO scientists often construct criterion situations on which to test an experimental program of instruction against the standard program. Since the objective of the Army testing groups was to construct the maneuver in a form that would "test" the ATFA field army against the standard field army, HumRRO scientists were presumably in a position to contribute experience and advice to the testing program.

The description of HumRRO's participation in building a large-scale criterion is designed to show how the large-scale criterion building differs from the smaller-scale efforts with which HumRRO is

---

1/ See Appendix A for further description.

familiar, and to outline the direction which the large-scale testing effort must take if it is to provide predictions better than the implicit estimates of experienced Army officers.

Building a large-scale "test criterion" for testing a field army is of course very different from building a "test criterion" for an infantry squad in the nighttime defense. One implication of this difference in scope, which should be carefully noted, is that in the small-scale situation the civilian scientist can absorb enough military information to become an "expert" in the small area on which he is concentrating. He can thus use a certain amount of implicit skill in abstracting the pertinent features of the situation which should be included in the criterion test.

The civilian scientist is not likely to absorb enough of the situation in which a field army operates to use the same implicit process of abstracting the important features for building a large-scale criterion test.

2. Methodology for the Abstractive Process

Behavioral scientists do not have a formal methodology for abstracting the important features of a job or task for inclusion in a training or selection program. Such scientists have for years used various abstractive processes such as task analysis, activity analysis, and item development, and on an intuitive, artistic basis they probably have some skill in this activity. However, the formal methodology for accomplishing these aims is only in its infant stages.

The process is the same as that involved in "transfer of training." The problem is one of selecting the smallest number of common elements between the criterion and training situations that give the greatest accuracy of predicting the accomplishment of one from the other. (The same abstractive process applies to the selection area, but the number of elements is generally smaller and the elements themselves simpler or more basic.)

It can thus be seen that the HumRRO adviser could advise on an "artistic" basis only. No formal methodology existed for prescribing the level of abstraction to be used in designing the test criterion, nor were any techniques available for abstracting the most important feature from either the combat or maneuver situations. Although advice regarding both these subjects was given and used, it is the thesis of this paper that continued advice on this artistic level will not contribute greatly to future Army tests of the type described here. Rather, it is held that if the contributions of behavioral scientists are to have value, the process of abstracting important features (isolating variables) must be attacked directly.

Many Army officers have advanced an alternative solution to the problem of task analysis: They suggest that officers be sent to school for a year or two in order to equip them with the skills for making such analyses. Since the problem of task analysis is so central to this paper further examination of this proposal is appropriate.

The parts of the college curriculum for scientific specialists are closely interrelated and interdependent; completion of the whole curriculum would be necessary before the officers would have the background they would need in order to attack their particular problems. On the other hand, it is possible for the Army to bring college professors to the specific situation under consideration. The advice the professor would give in this situation may be considered as the best lecture he can deliver on this specific subject; as he learned more of the situation his "lectures" would become more incisive. This seems more feasible than sending the officer to college, where his specific problem will not be dealt with.

It seems fairly certain that the large-scale Army testing program, unlike the small-scale test, cannot be handled by the civilian scientist, because he cannot analyse the field army in a few months with his present knowledge or tools. Since he cannot provide an answer in the form of an analysis of test criterion, some other solution needs to be devised. More time and effort with the present approach is recommended by the writer.

3. Implications for HumRRO

The implication of this problem for HumRRO is that civilian scientists are not familiar enough with large-scale Army operations to give appropriate advice for the construction of large-scale test-criterion situations without full-time attention to the problem. Continuous contact

with such operations would be needed so that scientists could either learn enough to isolate the pertinent features for inclusion in a criterion test situation or develop more formal procedures through which Army officers could do this.

In the opinion of the writer (and most Army officers) civilian scientists cannot learn all there is to know about field army operations, and it seems reasonable to assume that a piecemeal effort in the large-scale testing area cannot produce important improvements in testing programs. If this is true, only marginal improvements can accrue from HumRRO's efforts. Such activity reinforces the stereotype of the scientist as a person concerned with unimportant details, and the scientists involved likewise feel they are not working to good effect. This memorandum indicates the complexity of a large-scale testing program, and it is intended to serve as supportive evidence for the view that the problem cannot be effectively solved with a small, piecemeal effort.

II. The Problems of Organization and Doctrine Imposed by New Weapons and Equipment

A. Changes in Army Doctrine and Organization

For the first time in American history, technological advances in weapons and equipment have been continuing at an accelerated rate after cessation of actual war. Often the new developments so greatly improve military capabilities that adoption of new organizations and doctrine in order to utilize them is incumbent on the Army. The ATFA concepts represented one of those restructurings in organization and doctrine. Other changes are in the planning stage and appear likely to continue far into the future.

The nature of the alterations that should be made is not obvious. The Army must give up certain advantages for others when these changes are made, and whether the advantages given up are greater or lesser than the advantages gained is a matter of judgment in each case. Generally there is no quantification of the various outputs of the field army, so these judgments must be made through what we may call the artistic skill of experienced officers. It seems reasonable to assume that the changes in Army concepts will be of considerable magnitude in a few years; as the new weapons and equipment and their new supportive organizations and doctrines increase, the Army becomes, of course, less and less the Army with which experienced officers are familiar. The question of how far into the future experienced officers can extrapolate on the basis of their experience from 1945 and 1952 should be considered now.

Since it is one of the major concerns of science, the problem of extrapolation suggests a scientific approach. There are no a priori grounds for believing that the Army situation is so unusual that it would not be amenable to a scientific approach of isolating, quantifying, relating, and extrapolating. The only immediately apparent objection to such an approach is tradition--that the extrapolations have always been on an artistic basis and therefore they must be artistic in nature.

To examine the merits of this objection, let us consider a situation in which this view previously prevailed. By selecting an example where we have the advantage of the perspective that comes with hindsight we can judge the present case. Early pilots learned to fly without instruments--"by the seat of their pants," so to speak. When simple instruments were introduced in later aircraft older pilots often ridiculed these instruments and the younger pilots who used them; the older pilots generally could fly the new aircraft "by the seat of their pants" better than the younger pilots could with the crude instruments. This was an instance of formal quantification of a situation by use of quantifying instruments. The analysis of the situation needed for flying the aircraft could be made by the use of instruments or by an artistic analysis of undefined cues such as how hard the wind blew on the pilots face. We know that either procedure "worked," with the advantage going to the "seat of the pants" pilot for some time. Gradually the instruments became better and the aircraft grew so complex that the pilot could no longer get enough information from the "seat of his pants" to fly it.

We now have the advantage of perspective in judging this difference in flying technique. We know that in the limited frame of reference of the aircraft of his day, the older pilot was correct. But we also know that there is no question about the necessity for instrument flying in present-day aircraft. In this sense the older pilot was completely wrong. So it may be with present-day technological armies.

In most situations pertinent features can be abstracted and used to predict the whole situation. This is the process of isolating variables (or "main effects", in the analysis-of-variance parlance) and testing them for general applicability. Attempts to isolate variables in the new Army situation cannot be expected to yield astonishing findings for some time. Just as--for a time--the older pilot could do a better job than the young instrument flyer, the military expert can do a better job, implicitly, than can the first steps of an isolating, quantifying program. It is in fact an indication of validity when initial scientific approaches yield the same answers as experienced military judgment just as it was when the instruments became good enough to give the pilot the same information he was able to get previously by an artistic analysis of the situation.

B. Why a Test Is Desired

The average Army officer's conception of the scientific method is probably summarised in the word "test." Tests are generally used to

obtain information (at a given confidence level) regarding the desirability of alternatives; on the average, more confidence can be placed in the results of an empirical test than in the results of artistic, non-empirical analyses. Since the results of tests have the reputation of being correct in scientific circles, the Army often desires test results for their problems of restructuring.

C. Pitfalls for the Layman

Unfortunately the statements in the preceding paragraph are only generalizations that might be appropriate for the layman audience; the scientist knows that they are not true in all cases. "Tests" are not always used as a means of obtaining information; sometimes they are used as a device for convincing others (toothpaste ads are an extreme example of this). And if a test cannot be adequately controlled, it will not produce any better information than can be obtained through a non-empirical analysis. In fact, any test can be "rigged" by selection of inappropriate criteria. The layman is thus in a difficult position; he knows that tests are "good," but he cannot distinguish the conditions that produce valid data from those that do not. The scientist is not infallible in making these discriminations either, but he has some skill in the process and knows some weak spots to look for.

D. The Army Dilemma

Since they are not trained in the scientific discipline, Army officers must be considered laymen in the matter of empirical tests. The attitude they may be expected to maintain is that of any sophisticated

layman--that tests can produce better information than non-empirical analyses, but unless the good test can be distinguished from the poor, too much confidence may be placed in the wrong results.

The Army could do the tests itself so that it would know how to evaluate the results or it could allow civilian scientists to make the tests. There are real dangers in both of these approaches. The civilian scientist does not know enough about the over-all subject of Army warfare to ensure the proper construction of test criteria for large-scale tests. This is likely to produce biased conclusions to which an inordinate weight might be attached by the Army because the results are derived from "tests." On the other hand, tests conducted by military personnel only are not apt to be incisive. The abstractive process of <u>selecting the smallest number of elements needed for accurate prediction</u> is a crucial one in producing workable, valid tests. This skill is part of the general scientific method and is more likely to be exploited by scientists than by Army officers.

E. <u>A Possible Solution</u>

The familiar process of scientific validation appears to be a solution to the problem. The validating criterion in this case would not be actual combat but a "test criterion" which would represent combat (as described in section IV). By maintaining control over the elements in a test criterion, Army personnel could assure tests that were not biased because of the inexperience of civilian scientists in the subject of Army warfare. The scientific process could continue with the production of experimental test situations, and the criteria of these

would become more and more abstract from the test criterion. This process of abstraction would involve the isolation, quantification, relation, and extrapolation necessary for incisive, usable scientific predictions. But predictions would all have to be validated against this "test criterion" at intervals. Any failure of the experimental situations to predict the "test criterion" would indicate that something was wrong.

Of course, the abstractive process would not "belong" to the scientist; it could be accomplished only by teams of military men and scientists working together. The important point is that by the technique of maintaining "test criterion" situations the Army can ensure against results biased by civilian scientists' inexperience, and at the same time obtain the benefits of a scientific approach.

III. Scientific Advice During the SAGEBRUSH Maneuver

A. An Application of the Scientific Method in SAGEBRUSH

1. Advice Regarding Level of Abstraction

After the general discussion of the need for a scientific approach through the abstractive process, it may be desirable to provide an example of how advice of this type was offered to the AMTEG during SAGEBRUSH. This example will also serve as a referent for the more detailed analysis of the test criterion development, which will follow in the next section.

As mentioned previously, SAGEBRUSH was planned as a training maneuver only; the AMTEG had almost no control over the establishment of the maneuver conditions. When the AMTEG was established it consisted of only a few officers with no specific plans for conducting a test of the ATFA field army.

The AMTEG's initial approach was to ask the various technical schools and combat arms schools to submit requirements for the test. The requirements obtained by this procedure were generally in the form of questions, of the type that the Army uses in unit proficiency tests. It was only natural that such questions would be submitted, because the Army conducts this type of test every day of the year. However, the scope of the SAGEBRUSH test was much larger than any tests of a unit's proficiency; its concern was not proficiency <u>within</u> a unit but how well the units worked together. For instance, the concern was not with the tasks connected with vehicle maintenance, such as "check oil level," "tighten bolts," or "adjust cable," but with the output of the maintenance system, "how many vehicles were fixed each day."

The unit proficiency approach was not wrong on theoretical grounds; in theory, the detailed questions could be arranged to "add up" to the larger questions. If enough people had been available to go out and measure the details of performance within each unit the approach would have been appropriate, and such information would have been more "factual," less open to difference of opinion. On a field-army scale, however, this kind of approach would have required thousands of observers. The HumRRO adviser therefore advised against it and recommended that a more abstract unit of measurement be used. Here it is important to remember that explicit advice regarding a more desirable level could not be given by a civilian scientist without an extremely detailed Army background or any formal methods that were suitable for analyzing the task at hand. The adviser could only suggest a more abstract level and examine the product of the Army officers' implementation.

The AMTEG's next choice of abstractive level for measurement involved only five categories--mobility, flexibility, invulnerability, command control, and fire power--instead of the thousands of categories of the unit proficiency checks. The approach at this level of abstraction was almost decided on for the final units of measurement; however, some officers of the nucleus group of AMTEG found that they had difficulty in answering the questions they themselves wrote in these abstract terms. The problem was discussed with the HumRRO scientific adviser. He advised against the use of categories at such a high level of abstraction because

these variables had no usable definitions, operational or otherwise. The adviser felt that the use of such undefined variables would result in a mass of individual interpretations by field observers that would not "add up" to anything.

The decision was made to try another level of abstraction somewhere between the two rejected levels. The third level, finally used, was generally referred to as a systems approach rather than a unit approach, to distinguish it from the unit proficiency approach. "System" referred roughly to an organization of the type of the Signal Corps or Ordnance Corps, with a function independent of other systems rather than to battalions, companies, etc. The unit of measurement in this approach was in terms of the major output of each of the systems. An output was considered major if it went to a unit other than the one which produced it. The outputs which were directed wholly against the enemy were not generally measured, for certain administrative reasons which will be discussed later.

The outputs and inputs in this approach were called "packages." They were quantized by the joint conditions of who produced them and who consumed them. The term "package" referred to services (such as maintenance) as well as to concrete items. A package was not dependent on size or content--that is, amount or type of output. Nor did it depend exclusively on either the producer or consumer, but jointly on both.

2. How the Dependent Variables Were Measured

a. Adequacy of the package

Measuring the packages required a judgment by the evaluating officers. If the judgment had been made at a lower abstractive level (e.g., oil-gauge level) it would have been considered highly reliable--a "fact." If it had been at a higher level (e.g., mobility) it would have been highly speculative, because of the ambiguity of the category. At the package level it was somewhere between these extremes.

The measurement of the packages was in gross terms, because with the other large sources of error in the testing program it was not economical to attempt fine discriminations in judgment. The packages were judged in three degrees--adequate, inadequate, or marginal with respect to the inputs and outputs of corresponding systems in the standard army.

b. Techniques for partialing out confounding factors

One of the biggest problems confronting the AMTEG was the presence of contaminating factors.1/ The independent variable to be investigated was the field army organization and doctrine, and the dependent variable was the ability of the various component systems in the ATFA field army to produce outputs that were as adequate as those of the standard field army. The dependent variable was of course a function of numerous factors besides those under investigation, such as proficiency of

---

1/ A compilation of those factors prepared by the AMTEG may be found in Appendix B.

men and units, adequacy of equipment, scaling down of forces, and quality of umpiring. Under the maneuver conditions these important variables were allowed to vary in a random fashion with respect to the control group, the standard field army.

Since it was not administratively feasible to control these contaminating factors by holding them constant, the HumRRO adviser recommended an alternative technique for control--a technique designed to "partial out" the effect of the uncontrolled factors through the judgment of the observers of the action. This method was adopted for use in the SAGEBRUSH tests.

The persons who partialed out the effect of the contaminating factors were senior officers (major to colonel) called evaluators. The evaluators estimated what the value of the dependent variable would have been if such factors as proficiency and equipment had been the same as in a standard unit operating under the same conditions. After making this judgment, the evaluators were required to make an additional determination as to whether this adjusted value of the dependent variable was adequate, inadequate, or marginal with respect to a corresponding standard unit.

It was recognised that the "partialling out" technique made the data even more subject to personal factors and thus less reliable. However, the alternative--accepting the dependent variable as a representation of only organization and doctrine while such important factors as proficiency and equipment were ignored--was certain to produce erroneous conclusions. Losing reliability was felt to be less undesirable than

accepting a datum known to be incorrect. Only the most experienced officers served as evaluators, and they only after a day of instruction in the use of the method.[1]

3. Other Measuring Techniques

In addition to the evaluative technique two others were used. One was the comparative technique, in which there was no requirement to partial out the influence of contaminating factors. The officers who used this technique were required only to compare the adequacy of the actual performances with those of standard units in similar situations in combat. These officers were generally participants in the maneuver and received no special "schooling." They were sent questions on various aspects of the situation about which they were expected to be informed, and were furnished a page of instructions for making their judgments.

The third technique was a straightforward one of recording such data as times, frequencies, and temperatures. No special group was set up to provide this information; it was obtained primarily from standard formal reports and journals developed in the normal course of maneuver operations.

4. Relationship of the Various Measuring Techniques

It should be noted that even the more "factual" of the three techniques did not provide a check on the more unreliable techniques. For instance, if an evaluator judged a certain performance "adequate" and a comparative judge judged the same one inadequate both judgments had to

---

1/ This instruction was provided by the HumRRO adviser. (See Appendix C.)

be accepted as correct. In making his judgment, the evaluator may have considered the actual performance inadequate because of low proficiency, and may have felt that if proficiency had been more normal the performance would have been adequate. He might therefore enter "adequate" on his answer form. The fact that the comparative judge rated the actual performance "inadequate" had no relation to the evaluator's judgment of what it would have been if proficiency had been typical.

This point was not fully grasped by all members of the AMTEG. Many felt that the more factual data provided a check on the less reliable data. The HumRRO adviser made the point many times and on one occasion in a formal paper. (See Appendix 4.)

B. <u>Early Definition of Packages</u>

A procedure had to be devised for selecting and defining the outputs to be measured in the test. The HumRRO adviser recommended that the officers construct a list of the objectives for each system in the ATFA field army. These objectives were to be broken into subobjectives and sub-subobjectives until the latter were approximately the same level of abstraction. When some of the sub-subobjectives in one system were considered to be the appropriate level of abstraction they were used as models by officers working on the other systems. Help in determining the appropriate level was given informally by the HumRRO adviser. This procedure tended to ensure complete coverage of all outputs that it was appropriate to measure. Having the objectives at the higher level of abstraction made it easier to detect omissions at lower levels.

For each sub-subobjective one or more questions were prepared to determine the adequacy of output by the unit involved. (Objectives of the systems were outputs of the various subsystems or units of the larger systems.)

One of the early tendencies of the question writers was to "beg the question." An illustration of this tendency is such a question as the following: "Was command control more difficult when units were widely dispersed?" It was already known, of course, that command control would be more difficult, but the units had been dispersed to increase invulnerability. In this case, the question that really needed an answer would be whether adequate command control was possible under such conditions. The HumRRO adviser was active in pointing out instances of this kind.

The foregoing description is somewhat oversimplified; actually, many of the questions were irrelevant and/or did not follow from the objectives and their sub-categories or implications. Also, the level of abstraction varied from one question to another; some questions, such as the preceding example, were in terms of the discarded categories of mobility, command control, flexibility, invulnerability, and fire power. Although undesirable, this situation was manifestly one which would develop when the questions were being prepared before the approach to the problem was finalized. The AMTEG testing program was on a "crash" basis; to avoid in the future the sort of problem that arose, the HumRRO adviser recommended that testing programs be established four months before the operational planning.

C. Assessing the Reliability of Evaluator's Judgments

Generally, dependent variables were measured only once by Army officers. The number of evaluators assigned to ANTEC was not large enough to provide more than one measurement.

The HumRRO adviser found no way of assessing the reliability of the evaluator's judgments during the maneuver. Among other things, the same breadth and depth of Army experience that went into the making of the judgments would be required for checking their reliability. Also, the packages that were measured were usually not discrete. Most of them were rather continuous through time and their adequacy depended on amount per unit time as well as content. In many cases there was not enough time in the maneuver to be considered as a "unit time," especially in the resupply of materials that wear out rather than being consumed.

PART 2

SOME THEORETICAL ASPECTS OF MANEUVER TESTING

I. Introduction

This section contains suggestions regarding the nature of future tests of Army concepts. To take the suggestions out of the realm of pure opinion, certain theoretical points are developed before the specific suggestions are stated. These points will not be new to scientists; they have simply been selected and organized in such a way that they focus on the problem of tests of Army concepts. The writer feels that the suggestions follow naturally. If the reader disagrees with these suggestions it should be a simple matter to identify the basis of the disagreement in the theoretical analysis.

The applications of scientific principles to the SAGEBRUSH tests will be referred to or elaborated, to provide a basis for the more theoretical discussion as well as a demonstration of how these points apply to the relatively simple SAGEBRUSH tests.

II. Building a Test Criterion

A. Measurement of the Elements in a Test Criterion

1. The matching process

Whether a measurement or matching process is simple or complicated, it always requires the perception of a human being. For instance, to measure length, the scratches or marks on a measuring rod are matched or compared to marks on the edge on some other material. If the scratches on the rod match the marks on the other material, according to the perception of the observer, the second material is said to be the same length as the distance between the marks on the rod. There may be many "aids" to perception (such as optical systems), but even so, it is still a judgment of a human being as to whether a match has been accomplished.

The aspect of the measuring process that determines the "factualness" of the judgment is the amount of agreement from independent observers that might occur in a given situation. If all observers agree that a match has been accomplished the judgment is said to be a "fact." In the physical sciences measurement techniques are so advanced that we take short cuts in arriving at the appellation "fact"; for example, matching achieved by one person using an ordinary yardstick is usually reliable enough to be considered a "fact" for everyday usage. However, the very same measurement would not be considered a "fact" in a situation that required micrometer accuracy. Thus a "fact" is not absolute but always relative to the amount of accuracy required in a given situation.

2. Units of measurement

The matching process described requires that the comparison item have some aspect which can be measured in the same units as the criterion. For example, the tensile strength of a bar of steel is not directly measurable in terms of length, but we can structure a situation so that bar of steel will be responsible for an output that will give a measure of tensile strength in terms of length. When a heavy weight and a pointer are attached to the steel bar, the distance (or length) that the pointer moves is then an index of tensile strength.

This process involved the addition of two "systems," the steel-bar system and the weight system. The joint output of those two systems was in the same units as the criterion, namely length. By comparing the displacement of the pointer with the length criterion the tensile strength of the bar was determined. More exactly, the "tensile

strength" of the joint product of the two systems was measured. However, the weight and pointer system is held constant or "controlled" in measurements of tensile strength; the same weight could be used to measure the tensile strength of another steel bar. If the pointer moved a different distance with the other steel bar we would say that the latter had a different tensile strength. Of course, strictly speaking, the tensile strength would again be a joint product of the new bar system and the old weight and pointer system. But since the weight and pointer input into the joint system is identical in both cases we know that the difference results from the difference in the two bar systems. An analogy to the tensile-strength analysis of systems will be made later in this paper.

B. Selection of the Elements in a Test Criterion

1. Unimportant elements

Every system in an Army has certain outputs that are unimportant to the Army; by unimportant we mean that the system would function just as well without them. For instance, some outputs of systems that repair radio and radar equipment are burnt tubes, resistors, condensers. All outputs called "waste" are undesired outputs; there are many others that are neither desirable nor undesirable but merely of minor importance. It is often difficult to say what is of minor and what is of major importance, but such judgments are made every day; some things receive immediate detailed attention and others are relegated to secondary importance.

When a criterion is built for measuring a system or systems, only the outputs which are of some importance are selected for

measurement; all are not measured. Strictly speaking, the omission of these outputs from the criterion measurement is an instance of measuring from a reduced frame of reference. Whenever we measure in a criterion situation that is not actual combat we are leaving certain things out, but this need not be a disadvantage. We can make predictions from a reduced frame of reference and be accurate with respect to the important functions of a system.

It is still a matter of judgment, however, as to which outputs are important, or most important. Expending a given testing effort on the measurement of every conceivable output does not necessarily produce the most accurate prediction of effectiveness; it is more likely that more accurate predictions will result from expending more of the testing effort on the most important output(s). Of course, if an unlimited testing effort were possible, the best prediction would be made by measuring everything. The question of which things are the most important to measure is therefore a matter of practical, not theoretical, prediction accuracy. In the practical situation we almost always measure from a reduced frame of reference and get better predictions because of it.

2. Bias caused by selection of elements

Although efficient, the process of eliminating some outputs from consideration in a test is dangerous. It is the process that leads to bias in a test or enables a test to be "rigged" (that is, the results of any "test" may be predetermined by the experimenter by measuring only the "good" effects of a system and not the bad). For example, a toothpaste may be "tested" and found to destroy all bacteria in a single

brushing. This test may be rigged in that it may not include measurement of this toothpaste's effect on the lining of the mouth. The test may be quite factual about the effects on bacteria but fail to cover the effect that one brushing also destroys the lining of the mouth. Thus, it is not enough for a test to be "factual"; it must be unbiased as well.

Herein lies the danger of allowing civilian scientists to make tests and produce "answers" without a validation procedure controlled by the Army. In this paper the phrase "testing from a reduced frame of reference" refers to instances where such action has introduced a distinct <u>bias</u>, intentional or unintentional, into the test; it is not used in the strict sense, to mean that the slightest change from the actual to test situation constitutes a reduction in frame of reference.

3. Problem of reduced frame of reference in the Army

This practical matter points up a considerable problem in the Army. As a rule the higher-ranking officers have a wider perspective, or frame of reference, for the various outputs of the various systems than do the lower ranking officers. It is also true that an officer who is responsible for producing a given output may not be aware of the implications or importance of that output. His superior officer will have a better view of the output's implications but less knowledge of the details of producing it. The next officer up the chain of command has a still wider perspective and knows still less about the details of production. Thus the person who knows the most about a given system knows the least about how that system's output fits into the over-all frame of reference.

The point of this discussion may be summarized by a general rule: Outputs considered to be indices of the effectiveness of systems should be qualitatively and quantitatively defined by the person or agency with the appropriate perspective, and the measurements of those outputs should be made by the person or agency best qualified to measure them within the established definitions. This rule is not easy to apply when technological advances provide new outputs with which no one is experienced. Moreover, these new outputs might make obsolete large systems that previously produced similar outputs. (For instance certain systems with a nuclear output might conceivably make obsolete systems composed of certain conventional artillery weapons which produced non-nuclear outputs.) In such cases entirely new perspectives as well as new techniques of production will be required.

In a unit proficiency test there is little danger of testing from a reduced frame of reference, because such a test implies a situation where the responsible officer knows all the outputs and inputs of the sub-system. No new equipment, weapon, organization, or doctrine is being tested. An Army proficiency test is merely a test to see if the unit can produce its defined outputs the same way similar units have produced them in the past.

This type of test also implies a backlog of information on similar units which serves as a criterion for comparison. Presumably these standards of comparison have been obtained from combat experience; predictions of unit proficiency based on those conditions and standards

are likely to be accurate with respect to combat. In a proficiency test we want to know what is wrong or right inside the unit. However, these are exactly the conditions that will obtain less and less as the Army changes to utilize new technological developments.

4. Correspondence of outputs in the standard and "new" armies.

In the SAGEBRUSH tests it was not always possible to compare the outputs of a unit or system with the outputs of similar units or systems in the standard army; in some cases an exact one to one correspondence did not exist. For instance, in the ATFA army the maintenance functions of the Ordnance and Signal Corps were "split off" and formed into a new organization, a "maintenance system." This system had no counterpart in the standard army. In this case outputs produced by the maintenance system could generally be identified with the ones produced by the signal and ordnance systems in the standard army. These outputs were not identical, because they were not produced on the same schedule, location, or request basis, but there was enough similarity for the SAGEBRUSH tests.

It is also possible to go one step closer to the final field army outputs (outputs which are inputs to the enemy) and obtain measurements that were in the same units. That is, by measuring the outputs of units to which both the standard system and the new maintenance system supplied inputs, the effectiveness of the standard and new systems could be judged. In other words, the new and standard systems did not produce outputs that could be measured in the same units. However the

outputs of the systems into which the outputs of the new and standard maintenance systems went still produced outputs that were measurable in the same terms. Thus any difference found may be attributed to differences in the standard and new maintenance systems which made different inputs into the systems actually measured. (Naturally the systems actually measured must be equated by replication or experienced judgment.) This is one example of how changes in the Army will affect the details of given systems before they affect the outputs of systems that are closer to the final output of the field army.

5. SAGEBRUSH tests were not from a reduced frame of reference.

Several factors make it unlikely that the testing approach of measuring major outputs and inputs during SAGEBRUSH constituted a test from a reduced frame of reference. First, there is a background of knowledge from the standard army with regard to the outputs, even though unit structure may have been changed. Second, any input important enough to affect a unit's output would cause the officer responsible for that unit to call attention to the input as an important determiner of his output. This in effect would be a definition of importance. It would also be a derivative of previous experience with similar units. Third, the outputs to be measured during SAGEBRUSH (several thousand) were defined in advance by military experts through a technique of defining the major objectives of a system, the subobjectives of each objective, and so on. These hierarchies of objectives were reviewed for completeness by senior officers.

6. Smaller number of variables needed for future tests.

The earlier approach adopted by the AMTEG had been to measure everything in terms of five variables--mobility, flexibility, invulnerability, command control, and fire power--designed to describe an entire army. These categories are so abstract that they have not been defined in operational terms; they are much like intervening variables in a model. Any attempt to "measure" in terms of them would undoubtedly lead to confusion of meaning and omission of many important factors ("omission of important factors of outputs" is another way of saying "reduced frame of reference"). But this is not to say that that kind of approach is theoretically poor. It is probably good in the sense that construction of models with a limited number of the most important variables is a general scientific approach. This topic will be taken up later.

III. Sources of Error in the Tests Utilizing a "Test Criterion"

In the opinion of the writer, the source of error with which the psychologist is traditionally concerned--the perceptual error--is a relatively minor one in the "theory testing" experiments of the Army. The problem in testing new Army organizations will be one of selecting the most important sources of variation with respect to Army output. If only the small sources of variation are measured this will be a large source of error with respect to the predictive accuracy of the test.

The sources of error in the testing process may be categorized as follows:

(1) The perceptual error in matching.

(2) Limited control of confounding factors.

(3) The reduced frame of reference.

A. The Perceptual Error in Matching

Experiments imply measurement and measurement requires a standard against which to match the output of the system under investigation. In empirical studies a control group is typically supplied as the standard. Certain dimensions of output are selected for measurement on the standard and experimental groups, and these dimensions are measured and compared or "matched" for the two groups.

In studies of military subjects the control group is often a group from the standard army. It happens that there is a large body of knowledge about groups from the standard army. Army officers have seen these groups perform and know a considerable amount about their outputs. Officers are certainly not omniscient; in fact, for many purposes it is not exact enough to serve as a standard of comparison.[1] Still, the fact remains that certain things are known about such control groups, and this is somewhat novel in the experience of scientists who have done their work in laboratories. It would seem that for certain purposes, where gross measures are needed, information from experienced

---

[1] For instance, the number of casualties produced by a rifle squad at night is not known in quantitative terms from combat experience.

persons might be used as a standard of reference in lieu of a control group. This kind of "experienced opinion" control will not necessarily be inferior to the physical control group under all conditions.

The SAGEBRUSH study may be an example of the situation in question. Data regarding the output of units in a realistic setting were needed in SAGEBRUSH. In the strict sense, control groups would have consisted of units performing under the standard organization and doctrine (Aggressor side); however, the difficulties of making each control unit perform under the same conditions as its counterpart experimental unit (on the U. S. side) were practically insurmountable. The standard of reference was the expert opinion of officers who had seen similar units perform under many conditions in combat. It was not difficult for them to observe a SAGEBRUSH unit and compare its performance (or output) with that of standard units they had observed before. The dimensions on which the unit was compared did not require exact quantification but only non-parametric data--judgments of "better than" or "worse than."

The desirability of using control groups in such an instance is problematical. The error that would have resulted from "forcing" control groups into the same kinds of conditions as the experimental groups would quite likely have been greater than the errors in an experienced observer's recollections. It is argued here that the perceptual error is not always the greatest one in a practical situation, and that efforts to reduce that error may in some cases produce greater errors from other sources.

Studies in which no control group is employed should not categorically be considered non-valid. Since validity is a matter of

degree and dependent on accumulated error, each case should be considered on the basis of the experimental arrangement that results in the least accumulated error. (The umpires who assessed the units constituted a special case with respect to perceptual error, in that the evaluators' perception often had to follow from the umpires' perception and assessment.)

B. Limited Control of Confounding Factors

A second source of error in the testing process is allowing one or more of the systems contributing to a joint output to vary randomly while another system's contribution to that same joint output is being measured.1/ This is an instance of lack of control. Let us look at control this way: It can be achieved by control groups, but there are other techniques as well.

The output of any system is the product not only of that system but of all systems which supply an input to it. The output of a given system is thus the joint output of its own and all auxiliary systems.

1. The factor of external conditions

A special category of inputs are those introduced to a system through the external conditions in which the system performs. Weather and terrain are examples of this type of input in Army situations. The inputs of weather and terrain are from a system external to the Army; as important factors, however, they must be considered in a testing program.

1/ Factors which did vary randomly are listed in Appendix B.

2. The factor of inputs from ancillary systems

The inputs of ancillary systems do not need to be controlled by means of control groups; the necessary control may be achieved by artificial insertion of the inputs, that is by supplying the system under evaluation with "in tolerance" inputs from sources under the control of the experimenter rather than from other functioning systems. In that way systems may be tested one at a time. After they are tested individually it is desirable to test their joint output, because certain of the inputs introduced artificially might have been incorrect values or might have been from a reduced frame of reference. The over-all test might lead to further individual tests or the joint output might be satisfactory. The entire process is one of successive approximations.

3. The factor of internal conditions

Another source of error in tests of Army concepts of the same type lies in conditions internal to a system. If the system is not what it is purported to be, an error will result. In army systems this is a very real source of error. The proficiency of the men in one aspect of a system that may be out of tolerance. Condition and quantity of equipment is another. Still another is under-strength units, (under strength in men or equipment). This kind of uncontrolled variation was probably the greatest source of error in the SAGEBRUSH tests, greater than the errors of perception. Many units and/or systems were out of tolerance; in many cases the outputs of one system resulted in out-of-tolerance inputs to many systems. The communication system, for instance, was far below acceptable levels during the maneuver, and since this system

supplied inputs to almost all other systems the outputs of all those others were affected. This was also true of other systems besides communications, but to a lesser degree.

4. The combined error from all three factors

The over-all effect in SAGEBRUSH was extreme confounding. Weather and terrain inputs were fairly typical of the conditions for which the test was designed to predict, so this was probably not a large source of error. A large source of error lay in the internal structure of the systems; personnel were well below typical proficiency and units were under strength in personnel and equipment. And of course when one unit was out of tolerance for any reason it caused others to perform out of tolerance.

C. <u>Comparison of the Perceptual and Confounding Sources of Error in SAGEBRUSH</u>

There are no quantitative data regarding the amount of error contributed by the lack of control of these factors as opposed to that contributed by the perceptual comparisons in the SAGEBRUSH tests. The writer and all the members of the AMTEC believe that the contaminating influence of out-of-tolerance systems was by far the greatest invalidating factor in the SAGEBRUSH tests. If tests of this type are to be made more valid, future scientific and military effort should be directed toward this source of error rather than toward the perceptual source of error.

D. <u>The Frame of Reference Error</u>

One other source of error exists: Error is produced by measuring from a reduced frame of reference, a matter already discussed

in general terms. In the SAGEBRUSH tests this was probably not a relatively large source of error. An attempt was made to have as many systems operating as possible, and, with the AMTEG procedure of identifying objectives and subobjectives for review by all Army schools, probably no major outputs of these various systems that were inputs to other systems were overlooked.

For administrative reasons, the outputs that were inputs to the enemy forces were not all considered for measurement. The AMTEG was given responsibility for testing only on the U. S. side; not until after some work on the problem of testing had been done was it realised that the over-all effectiveness of the U. S. side could be measured by its effect on the Aggressor side. Rather than change plans at a late date, it was decided that the effect on the Aggressor side would not be measured.

Certain other practical considerations also entered into this decision. For one thing, the joint output of all the army systems would be contaminated by so many sources of error that it would be almost impossible to partial out the confounding factors from the factors of organisation and doctrine.

Testing of this over-all output is certainly a goal of future testing programs. The AMTEG felt that under the conditions of the SAGEBRUSH test such measurement would not contribute enough to warrant a change of plans a month or so before the test was to be finalised.

Testing from a reduced frame of reference thus appears to have been a relatively small source of error in the SAGEBRUSH tests. The largest source of error remains the lack of control of inputs from the various systems.

IV. Continuum of "Test Criterion" Complexity

Let us now carry our analogy of tensile-strength testing into a military setting. The following examples are designed to show that as we test variables with broader and broader implications we extend the necessary testing program far beyond our present measuring capacity.

A. A Simple Test

Let us start with a simple test--evaluating a newly developed shoe sole. It is entirely within our present capability to validly test a new shoe sole. We can add the new shoe-sole system to an abrasive system and measure the time it takes to wear through. The same abrasive system can be applied to the old shoe-sole system and a comparison made. The abrasive system need not be a highly structured one in a laboratory; it can be men wearing shoes.

It may be seen, however, that some control is lost when we move out of the laboratory situation; that is, we are not so sure that the abrasive system is identical for the two shoe-sole systems being compared. We cannot be sure that the two sets of shoes will get the same treatment. On the other hand, we are more certain that the types of treatment which cause abrasion in field use will occur in a "troop test" than in a laboratory. This example illustrates that we have adequate techniques and

instruments (both laboratory and field) for validly testing simple things such as shoe soles. We have used shoe soles for a long time and have accurate criteria for their wear.

B. A More Complex Test

As a somewhat more complex example, let us take a weapon such as the rifle. The criterion for a rifle's effectiveness is in terms of the number of casualties it produces in enemy troops. This criterion is available through the judgment of experienced officers who have observed the effects of rifles in combat.

In order to obtain a rifle output in these terms a number of systems must be added to the rifle system. To the system involving the men needed to fire the rifle must be added the conditions (such as terrain, weather, light) under which it must be fired. Enemy troops must be added as a third system; other factors such as weight and numbers of cartridges required could be added as still another system involving supply. We can test the output of a given rifle system against another in terms of the criterion of enemy casualties produced when these other systems are added to make a joint output and are kept constant over different rifle systems.

Keeping the output of all those systems constant is, however, not a simple matter. The more systems that are contributing to the joint output, the greater the number that have to be controlled in order to measure the effects of one of them. From a testing point of view it is much simpler to eliminate other systems from a test than to hold them constant. This procedure is desirable if a criterion of the effectiveness,

in isolation, of the system under study is available. Such a criterion is not available for the rifle system alone, because in combat no one ever sees it in isolation and no effort has been made to obtain data for such a criterion in a controlled situation. Obtaining such data is a step that can be taken only after the effects of the system under study are quantified and functionally related to the joint output of the contributing systems. Therefore, at present we only have information regarding the joint effect of the rifle and other systems. Study of rifle effectiveness should proceed with that criterion. The problem of controlling the other systems contributing to the output, which is in the same terms as the criterion, must be accepted and solved if the studies are to be valid.

In considering the example of the rifle it might be noted that the accuracy of a single slug at extreme ranges has often been used as a criterion of the combat effectiveness of rifles. By eliminating all other systems, this criterion makes tests quite simple. Unfortunately, however, the functional relationship between rifle accuracy and the combat effectiveness of rifles is not known (and it is certainly hazardous to assume they are identical). Therefore, basing predictions of the combat effectiveness of rifles on the accuracy of the rifles will not result in the best predictions.

The rifle was included here as an example of a system where implications are so broad that it is beyond our present capacity to validly test.

C. A Very Complex Test

The next example is one in which the implications are even broader: the tactical atomic weapons now becoming available. We have no experimental criteria regarding the effect of these weapons in combat, but we do have some knowledge of their effects from the proving ground tests. From that knowledge we know that they have the same type of effect as the existing weapons systems (rifles, tanks, artillery, etc.), but we know also that their potential is greater. Possibly the effects of these other weapons systems may be better achieved by the atomic weapons systems than with the system currently organised for that purpose.1/

In order to solve the problem of what balance of weapons systems will achieve the greatest over-all effect, a criterion must be available. This criterion must be in the same terms (or units) as the sum of the terms of the various systems involved. Such terms are almost identical to the terms in which actual combat is constituted. In addition to the weapons system, the systems of intelligence, communications, command structure, and conditions of terrain and weather are all involved. And each of these may be broken into sub-systems and sub-sub-systems, the smallest of which would be on a par with the factors involved in the wearing out of shoes.

1/ This example is not meant to raise the argument that every time a new weapon is introduced someone thinks that it makes all old weapons obsolete. The example is included for another reason and a discussion of this point is irrelevant to that reason.

Assume for the moment that such an appropriate criterion can be constructed. The problem of controlling the effects of all these systems in the joint product while measuring the effect of the atomic system is one of considerable proportions. However, this question requires a valid answer. The atomic weapons example is included as an extreme case having implications for many systems; as the implications grow, so does the difficulty of controlling the outputs from all of the systems. A maneuver such as SAGEBRUSH is an attempt to produce an appropriate criterion, but the control of all the various systems is an imposing task. The large number of "contamination factors" present in SAGEBRUSH attests to the inadequacy of our present capacity for tests on this scale.

D. Why We Are Concerned With Joint Outputs

The three examples, shoe soles, rifles, and atomic weapons, were selected as three points on a continuum--that of an increasing number of systems involved in producing a joint output.

We must deal with this continuum because we are interested in these joint outputs and because we usually know something about the criterion situation for such outputs. We are interested not in the rifle, per se, but in the joint effect of a number of systems of which it is only one. We are not interested in the inherent accuracy of the rifle, since we do not know the functional relationship between its accuracy and its combat effectiveness. Therefore, because of the nature of our interest we must deal with a fairly large number of component systems.

There is nothing inherent in the various systems themselves that causes us to be concerned with their joint effect; it is simply that in the over-all activity of war we are interested in producing casualties. Corollary to this statement is the fact that our experiences with criteria are in terms of the things we are interested in. That is, we have some knowledge of the number of casualties that can be produced by rifle fire because we are interested in that effect and have observed it. We must therefore test those systems that produce joint outputs which are in the same terms (units) as the criteria which we know something about.

We do not now have the capacity for adequately testing any but the simplest of systems (as is currently done by CONARC boards). There is no reason, however, to believe that the process of quantification and determination of functional relations cannot be applied to military situations just as it has been to other subject matters. The first step is quantification; the second, development of functional relations.

V. Predictions by Quantification and Functional Relationships

A qualitative approach to systematization is conceptually simpler than a quantitative one and generally comes before quantitative adjustments in science. In general, from the standpoint of test development, the Army is in the first stages--the qualitative phase. A "good" army, in this scheme, should have certain qualities:

(1) Flexibility

(2) Mobility

(3) Invulnerability

(4) Command control

(5) Fire power or shock action

In the qualitative conception an army is "good" if it has these qualities and "bad" if it does not. This approach is certainly not wrong, but let us examine its power of prediction. As an example, consider a Roman legion. Undoubtedly an army of today could easily defeat a Roman legion; yet if the legion is examined for the five qualities, it will be seen to have had them all. Its invulnerability was achieved by shields and body armor, its fire power by spears, swords, and arrows, its mobility by marching, its command control by voice commands or simple signals, and its flexibility by whatever SOP's were used to redistribute the effort of their special equipment. Thus it may be seen that a <u>qualitative</u> analysis is not a sufficiently incisive tool to describe differences between a modern army and a Roman legion that can be used to predict the winner of an engagement between the two.

However, if each quality is assessed in <u>quantitative</u> terms, the Roman legion is shown to be different in terms of our "theoretical analysis," just as we know it would be in practice. Today's army has greater fire power; in certain respects, though not all, it has greater mobility and more invulnerability. The advantage that the modern army has in command control requires a finer quantitative discrimination--although control is augmented by electrical equipment, the modern army is more dispersed and hence out of range for direct control by voice. Thus it may be seen that a quantitative approach has more predictive power than a strictly qualitative one.

One more step can be taken for greater predictive power: The parameters of the interrelations between qualities can be established. Basic to this discussion is the fact that the five qualities are

interrelated. For example, a certain amount of invulnerability may be achieved by dispersion, but dispersion in turn causes a certain loss in command control. The question may be posed: How much command control should be given up for so much invulnerability? Such a question not only assumes quantitative measures of both command control and invulnerability but also requires some knowledge of their interrelation. Knowledge of this interrelation (in the form of a curve or function) would lead to much greater predictive accuracy than would simple quantification. In fact, if the interrelations of all the qualities were known quantitatively, the theoretical formulas would have extremely practical predictive power. The accomplishment of this quantification and formulation would be the objective of the scientist.

The five quality categories should be considered only as examples or points of departure; they are not necessarily the best ones to use in the proposed formulation. (Perhaps the 10 principles of war would be better.) The skills of both the military men and the scientists are required in order to structure the categories or qualities that would be most easy to handle and yet have sufficient predictive power. The categories would have to be general enough to allow future equipment and weapons to fit under their rubrics, and categories which lend themselves to quantification would of course be more effective than ones which did not.

The approach outlined is certainly not new in science; in fact, it is typical. It is stated only because the writer has not seen it spelled out in writing for large-scale tests of Army concepts. Research on a

smaller scale can generally be done on a more rote basis; for example, once a criterion is established for day or night firing of the rifle, a training program can be constructed and tested against the existing program. This is a discrete solution; it does not necessarily involve establishing the dimension of relationship between the various possible programs. In the proposed approach qualities, categories, or dimensions must be established, because of the great number of factors which affect the final product and the need to allow for new conditions (equipment and weapons).

The proposed approach does not necessarily fall entirely within the purview of HumRRO. Certainly the capabilities of humans (including such factors as capacity for training leaders) are an important consideration in establishing the various parameters. The general approach of quantifying and formulating may lie outside HumRRO's responsibility. But the question might arise, "Whose concern is it?"

VI. Future Tests of Army Concepts

A. The Approach to Future Tests

New equipment and weapons are being developed at a rapid rate. Some weapons have outputs similar to those of standard army units or systems; others have outputs which the standard army could never produce previously. New developments in equipment make obsolete some old types of equipment, and in other cases make possible things that could not previously be done at all. If one considered only the weapons and equipment that affected outputs which are, to some degree, in the standard army's repertory, the changes in army structure (organization

and doctrine) that would be required to most effectively utilize them would not be great. However, new weapons and equipment can produce outputs that cannot be achieved even to a degree in the standard army. The integration of these new outputs will have far-reaching implications for Army organization and doctrine.

As long as the changes in structure are small there will be a correspondence between the standard and the new Army units. Under these conditions experienced officers can compare the outputs of units in the new system with what they know of outputs of similar units in the standard systems. This process involves a certain amount of extrapolation, but as previously stated, this is not necessarily the largest source of error in a test of this type. However, as changes in structure require units that are less and less like the standard units, these extrapolations will become larger, more difficult, and subject to greater error.

The changes will probably come first in the smaller units rather than in larger organizations. For example, new capabilities of radio equipment will cause changes in the smaller units that work directly with telephone and radio equipment, rather than in the Signal Corps. And if atomic shells for 105 howitzers became available in quantity, changes in the organization of howitzer batteries might be required (assuming one atomic weapon could produce a saturation casualty effect in an area that previously required continual fire by several howitzers to produce saturation). Such capabilities might in turn produce new equipment requirements.

The over-all output of the field army will be the last to change completely. That is, the dimensions by which the over-all output can be described will be appropriate long after the dimensions specific to subunits are outdated. However, the quantities, or values, on these over-all dimensions will change; with smaller units the dimensions themselves will change. This may be seen better in terms of levels of abstraction. The descriptive term "animal" is more abstract than "horse", "cow," "duck," or "pig." Certain predictions about the behavior of all members of the animal kingdom can be made on the basis of the term "animal," and those predictions will be useful. For instance, we may predict "movement" for animals. This prediction will be correct for a group of pigs, horses, and cows as well as for a group of chickens, ducks, and geese. Of course, by being more specific we make more accurate predictions; predictions of amount of movement will be more exact if we use the categories of quadruped and fowl, rather than animal. Both kinds of predictions have their place, however.

In the military situation, an army can be described in more abstract terms than those used to describe its component units and/or systems. The useful feature of the more highly abstract descriptions is that predictions can still be made even if the components change.

The following scientific procedure should be considered as a systematic procedure for aiding the higher-level officer in specifying the required outputs of a field army. By analysis, the outputs that supply the best predictions of over-all effectiveness should be isolated. The categories should be general enough so that new weapons and equipment can

be subsumed under the same rubrics as the old, although with different values of output. This implies that the categories would be strustured in terms of outputs. The objective of the procedure is to devise general categories that do not need to be changed qualitatively in order to insert new developments. The categories should be such that new weapons and equipment developments can be assigned new quantitative values in the old units of measurement. The functional relationships established with the old values of these general categories will presumably hold with the new developments. Predictions can therefore be made by entering the new values in the functional relationships and then extrapolating to the new values of the other functionally related categories that will obtain with the new value that is entered.

To illustrate this procedure, let us consider the categories of mobility, flexibility, invulnerability, command control, and fire power--highly abstract terms currently used to describe the Army. There are functional relationships between these categories, but what the relationships are in quantitative terms, no one knows. Estimates of the quantities are a matter of artistic judgment at the present time. However, it does not seem unreasonable to believe that the five terms can be quantified, and after quantification they could be functionally related to each other in useful quantitative terms. Once these quantifications and functional relationships were developed, extrapolations to situations involving new equipment and weapons could be made with accuracy and assurance rather than on an artistic basis. It is argued here that as

the Army changes more and more from the standard Army with which experienced officers are familiar, their artistic extrapolations from combat will have to be greater and consequently more subject to error. Again, it is not argued that the five categories discussed are the most useful dimensions for describing the Army; they were selected simply as examples. They are currently in use by Army officers, though not in a quantitative or functionally related sense.

The point is that the Army plans for reorganisations necessitated by new weapons and equipment can be approached scientifically. Further, it is argued that such an approach deals with the Army as a whole rather than simply with aspects such as training, selection, and operations. These aspects fit within the larger scientific framework and presumably all of them are structured by the over-all picture.

The effect of the over-all approach suggested here would probably be to provide a certain kind of continuity for the changes which will be made during future years. We might say that the dimensions of change will be identified.

There is no reason why these dimensions cannot be put into language for Army consumption. Knowing the dimensions of change might have an extremely tangible and salutary effect on most Army operations. One effect might be that of orienting Army personnel at all echelons to "where they are going," how their job will change, etc.; such knowledge would possibly remove the disruptive personal aspect from the changes. Subjectively, the process would be a continuous one rather than the "crash change" process it is now and will be with further changes. This kind of benefit would be added to the effects on training, selection, procurement, planning, and other operations.

In some ways this section would seem to negate some of the points suggested earlier. That is, it was previously suggested that the recollections of experienced Army personnel could serve as a control under certain circumstances, and also that the measurement of the final over-all output of the field army was highly contaminated. Actually there is no conflict in either case. They both involve the time factor. At present, experienced personnel can make certain comparisons of the standard and "new" armies, but in the future they will not be able to make those same comparisons because the new units will not correspond to the old. The case of measuring the final product of the field army is similar. In SAGEBRUSH so many factors were confounded that it was impossible to untangle them in the output of the field army; however, this need not always be the case. In fact, it is reasoned here that the contaminating factors will have to be controlled, because we cannot continue to measure and compare outputs from subunits as these units lose their identity in the new organizations.

In effect these points represent a recognition that officer judgments constitute good prediction at the present time, but that they will lose predictive power in the future; a program based on tests of the over-all output of the field army will have to be developed before accurate predictions can be made in the future.

B. Future Test Vehicles

Will tests of future organization and doctrine use maneuvers such as SAGEBRUSH as a vehicle? It seems to the writer that maneuvers cannot do the job; at any rate, not by themselves. Extensive testing of

sub-systems and/or units is required to "shake them down" before a maneuver, but aside from this, a maneuver has inherent disadvantages. Most of these can be covered in the statement that maneuvers cost too much in terms of time, money, and disruption of other activities.

Nothing has been said about the N of SAGEBRUSH test. But it was one. Even if there had been no contaminating factors and all measures had been perfect, the N would still be one. SAGEBRUSH was conducted in a given terrain, weather, and roadnet situation. And most important, the split of logistic support from tactical command depended in great measure on the reactions and personalities of the specific commanders involved for its success or failure. Even with the best test of this particular set of commanders, the question of how well it would work with other commanders would remain problematical. However, the expense of obtaining a sufficient N would probably be prohibitive. These points underline once again the necessity of developing predictive tests which have fewer elements than the test criterion.

The abstractive work of scientists and officers might be facilitated by the use of electronic computers; these would certainly solve the problem of N. Parameters for the machine could be collected from maneuvers, combat, war games, and CPX's. New organizations of the Army could be "played" by commanders and their staffs, with machines rather than actual troops used as subunits. Machine "wars" could be played on two sides. The same "war" could be fought with different commanders, different terrain, different weather, and what is most important,

different field-army structures. Electronic computers may not be the answer to the N problem, but there is a need for an abstract test that can be repeated under controlled conditions at a reasonable expense.

Maneuvers could be used as a validity check on the results of electronic war gaming or other abstract tests that might be developed. This would provide a test criterion in which the Army could maintain control of the frame of reference.

Enough data would be obtained from the electronic war games to describe the functional relationships (curves) for weather, terrain, roadnet, and all those factors which will be constant in a given maneuver. Predictions could be made from the electronic data for a given set of conditions, chosen at random; those conditions could then be duplicated for a maneuver and the electronic predictions validated on the maneuver outcome. The accuracy of this prediction would serve as an index of validity for all the functional relationships developed by the electronic war games.

PART 3

HUMRRO PARTICIPATION IN FUTURE ARMY TESTS

I. Probability of The Inclusion of Tests in Future Maneuvers

Some kind of empirical testing will probably be desired for the Army's new theoretical concepts of reorganization. This judgment is based on the opinion that the results of the BLUEBOLT, FOLLOW ME, and SAGEBRUSH tests will produce enough new information on the ATFA concepts for CONARC to consider them to be of value. Also, since "test" generally has a positive valence, "test" results, will be desired in the future. Not enough is known about such large-scale "tests" as SAGEBRUSH to permit a correct evaluation of their validity. As long as there are no well-known ways of distinguishing degrees of validity for these tests any kind of test will be considered adequate. Finally, a precedent exists for testing concepts like ATFA, which restructure the Army, by means of maneuvers. This procedure is likely to continue until other procedures are demonstrated to be better.

The nature of the next maneuver test of doctrine and organization will be similar to that of the SAGEBRUSH test unless a scientific advisory group suggests another approach. If the procedures used by the AMTEG in SAGEBRUSH provide satisfactory information the same procedure will be followed again.

II. Assuming That HumRRO Does Not Participate

The following predictions are based on the assumption that HumRRO will not participate in future maneuver tests. It is predicted that, in the absence of a scientific adviser from HumRRO, the AMTEG testing procedures will not be used in the way they were designed to be used. In this event the testing agency will be dissatisfied with the procedures

and will recommend new procedures for the next maneuver. The recommendation will very likely take the line that testing with questionnaires, objectives, and subobjectives is a waste of time, and that experienced officers do not need all this to do their job. The testing agency will probably recommend that more officers serve as evaluators and that fewer "long hair" techniques be employed. This recommendation will probably be followed, which will put the situation back where it was with the test of the Triangular Division in 1939. The cycle will then probably repeat itself.

If HumRRO were to continue participation by the assignment of a scientific adviser for brief periods, the type of test conducted for SAGEBRUSH would probably continue as long as he was so assigned. In the opinion of the writer, the adviser could do no more than sustain the present procedures in the amount of time he would be able to devote to the job. Considered in terms of the writer's thesis that the present procedures will become ineffectual as the Army changes more and more from its current structure, such action by HumRRO would not contribute much to the effectiveness of the tests.

The implications of greater degrees of HumRRO participation are obvious. In the writer's view any participation should follow the direction described in this report. Following such a course would be desirable from the standpoint of developing effective tests for future reorganization, but it would, of course, lead to other problems. For example, the question of whether this is a legitimate HumRRO activity would arise as would the question of how much effort can be expended.

The writer is not attempting to answer these questions; in developing his ideas of what a testing program will have to be to be effective in the future he gave no consideration to whether HumRRO had cognizance or interest in the area.

III. Advantages to HumRRO Participation

The theories of organization and doctrine envisioned for the Army of the future imply rather wide departures from the present Army. These changes imply that activities of individuals and groups will differ markedly, and under these circumstances individuals must presumably be trained differently. For example, the wider dispersion that will characterize the armies of the atomic age means that patrols will be able to penetrate deeper and more often into the area occupied by dispersed enemy troops. Patrol action over wider areas and in penetration of several miles will mean that the patrol members will not personally observe their routes before the patrol begins. This in turn will mean a greater training emphasis on map navigation than on navigation by previous observation.

All such differences in activity, and they are probably legion, point to modifications in training. Bigger training modifications will result directly from the introduction of new equipment and weapons. The training programs for new weapons and equipment are not generally in existence before the "hardware" is available in quantity, and the possibility also exists that new equipment and weapons will make certain current activities obsolete. It seems to be to HumRRO's advantage to know what changes will occur, and to tailor its training research accordingly.

Information on projected changes will not be available in concrete form, since a certain amount of analysis of the situation is required in order to obtain it. In the writer's opinion, information of this kind will be available only when human-behavior scientists abstract it during a maneuver, although some of the more obvious differences could probably be ascertained from analysis of the concepts themselves.

The important point is that prior to some kind of analysis this information is not in a usable form. In order to learn the subject matter which they can then analyze, HumRRO scientists should work with the Army during maneuvers.

Three points have been made in this section: (1) future changes in the Army will be manifest in maneuvers, (2) these changes have important training implications for HumRRO, and (3) to obtain the training information, HumRRO scientists must perform on-the-spot analyses of maneuvers and of concepts as well.

Such analyses would benefit the current programs in which HumRRO engages; the activity of planning Army theory testing programs is another matter. The question of whether direct benefits would accrue from participating in such an area should be considered separately.